



Pęknięcie czarnych skrzynek

Często, gdy ludzie mówią o matematyce, słyszysz słowo algorytm - proces wykonania określonego zadania, krok po kroku. Na przykład kroki, jakie podejmujesz podczas dzielenia lub mnożenia ołówkiem na papierze tworzą dwa bardzo dobrze znane algorytmy. Algorytmy mogą być bardzo użyteczne, ale ostatnimi czasy, przy tak wielkiej ilości tworzonych i udostępnianych danych oraz wraz ze wzrostem ich wykorzystania w krytycznych obszarach, takich jak zatrudnianie, udzielanie kredytów czy opieka zdrowotna, algorytmy są poddawane intensywnej kontroli pod kątem ich bezstronności. Ludzie doświadczają skutków konsekwencji płynących z zastosowania algorytmu, ale dane i kroki, które stanowią podstawy do takiego wnioskowania są często przed nimi ukrywane (jak wewnątrz czarnej skrzynki). Matematycy i wielu innych domagają się większej przejrzystości i odpowiedzialności, aby algorytmy mogły zostać zbadane w celu określenia, czy lub w jakim stopniu wykazują one tendencje stronnicze na różne sposoby, w tym uprzedzenia ze względu na rasę, płeć, wiek i pochodzenie etniczne.



Badacze proponują kilka sposobów zapewnienia bezstronności. Jedno z całościowych podejść wymagałoby od projektantów algorytmu udowodnienia, że ich algorytmy są bezstronne jeszcze przed ich wdrożeniem i to zamiast oczekiwanie, że użytkownicy sami odkryją uprzedzenia. Innym sposobem byłoby wykorzystanie technik zapożyczonych z socjologii, które mogą ujawnić stronniczość, na przykład poprzez odpowiadanie na pytania typu „Jeśli ta osoba byłaby innej rasy, czy zostałaby zatrudniona?” Jakakolwiek metoda zostanie użyta lub jakiegokolwiek przepisy przyjęte, celem badaczy w tej dziedzinie są algorytmy, które są przejrzyste tak, aby uzasadnienia podawane dla ważnych decyzji przestały brzmieć „ponieważ sztuczna inteligencja (AI) tak powiedziała”.

Tłumaczenie: Magdalena Migda, dzięki uprzejmości Polskiego Towarzystwa Matematycznego.

Więcej informacji: *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Cathy O’Neil, 2016.